5.14 Exercise: Explore sampling variation - proportions

This exercise will enable you to use Visual Inferential Tools (VIT) software to look at sampling errors of proportions. VIT will enable you to see the variation between samples and illustrate how the variation changes with different sample sizes.

VIT is a developing collection of software modules for use in exploring the core concepts involved in drawing conclusions from data.

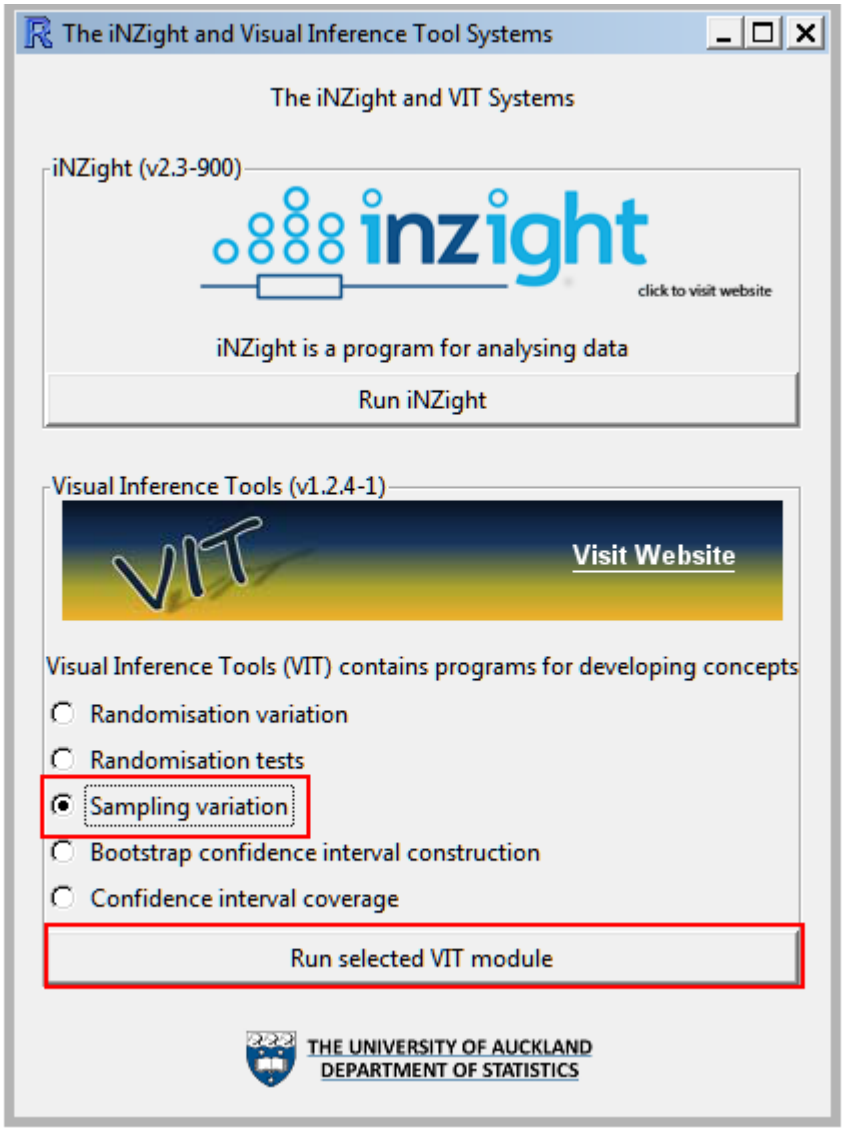
The skill addressed is:
Learn how to use VIT Sampling Variation to produce a sampling distribution for proportions like those used in the Random Error video.

[*VITonline version linked here*]

INSTRUCTIONS

Follow the instructions below to generate the graphs. Or you may prefer to print these instructions.

Start iNZight as instructed in the previous Exercise and select the **Sampling variation** option and click **Run selected VIT module**.



In the previous Exercise you saved the **NHANES-1000** dataset on your desktop. Use **Import Data** and the browser window it opens up to import the **NHANES-1000** dataset into VIT from your Desktop.

Click **View Variables** to see a list of the NHANES-1000 variable names.

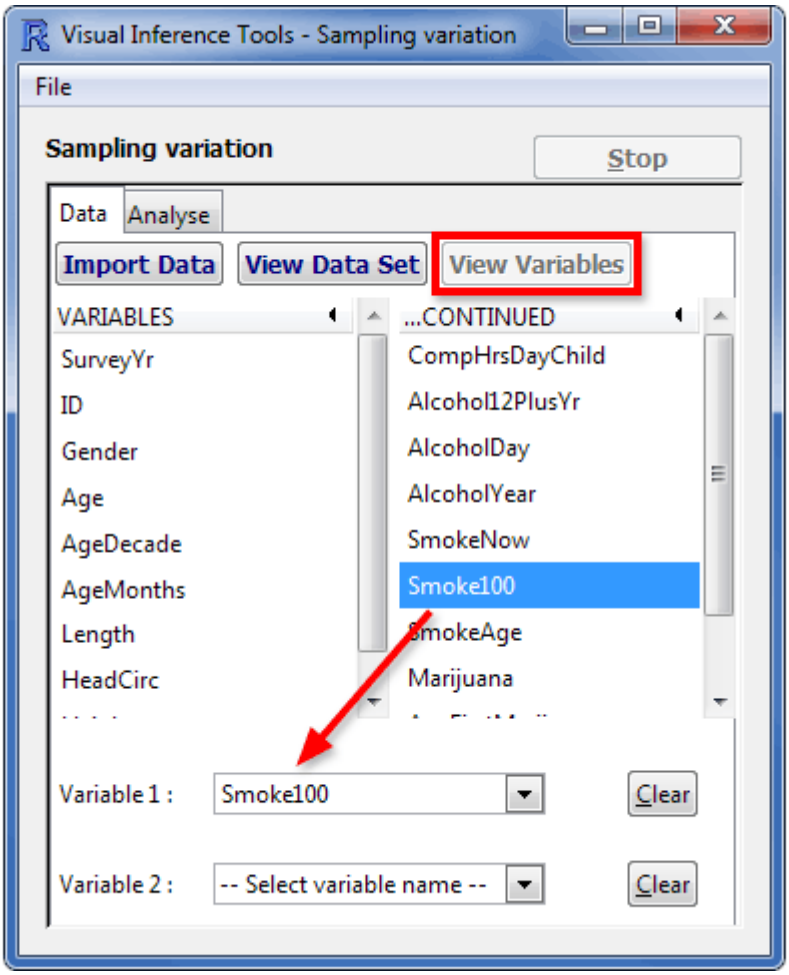
Under the **Data tab**, scroll through the variables and remind yourself which variables are categorical and which variables are numeric. In this exercise, we are looking at categorical variables.

If you have any problems during this exercise, see **Common questions** at the bottom of this page.

Using iNZight VIT to generate a sample proportion

We will explore the variable **Smoke100** which is the proportion of people in the NHANES-1000 dataset who have smoked more than 100 cigarettes in their lifetime.

Drag Smoke100 down into the **Variable 1** slot. You will see the proportion of people who answered 'yes' or 'no' to the question "have you smoked 100 cigarettes or more in your lifetime?"

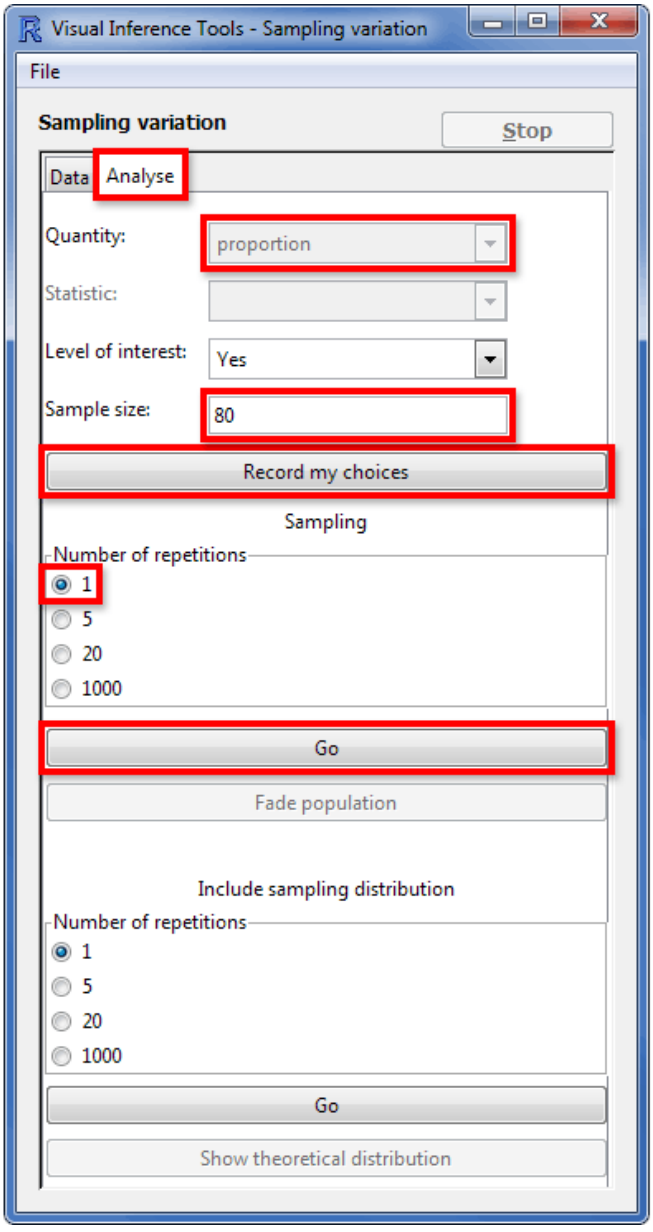


Now click the **Analyse tab**.

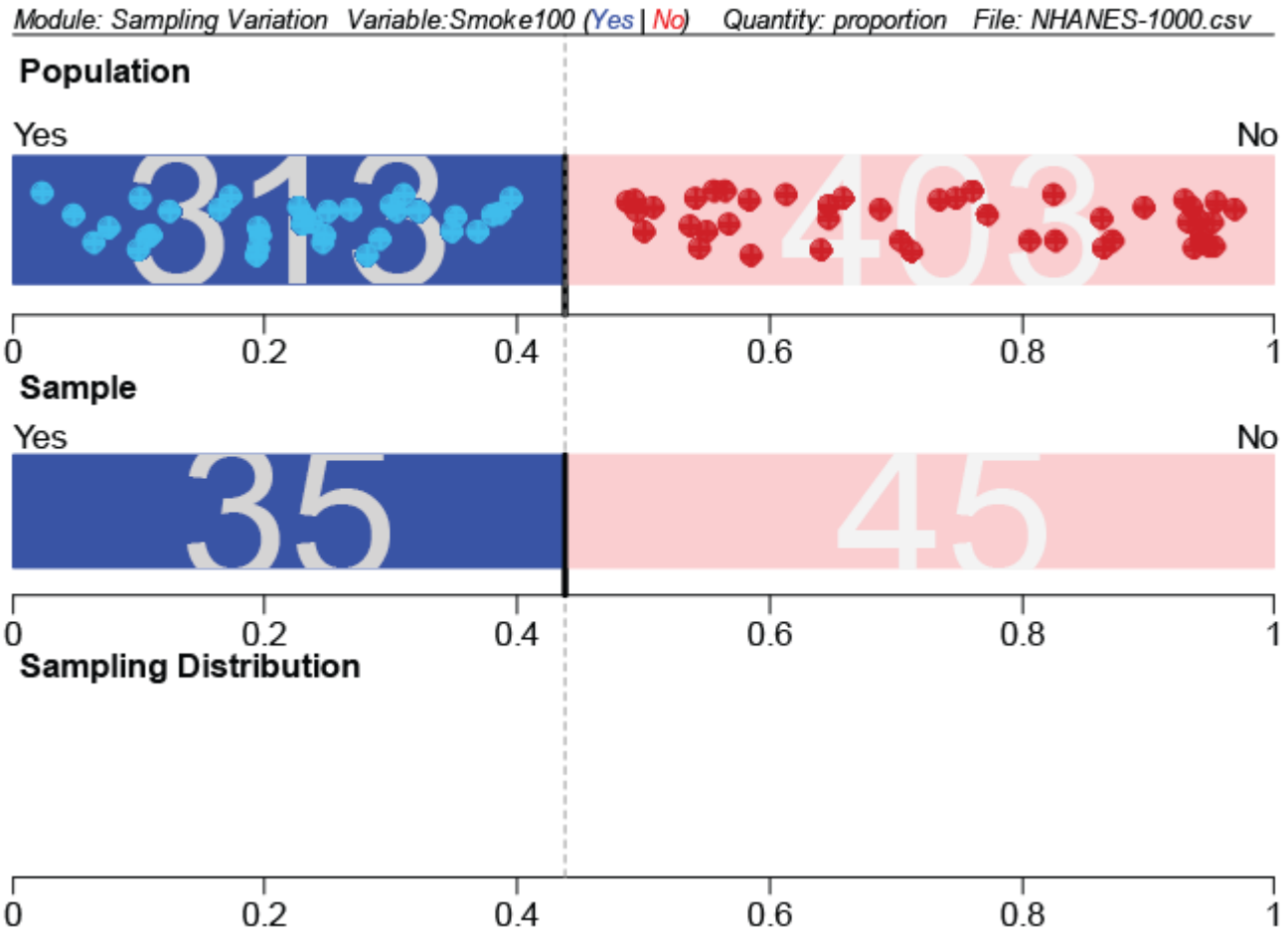
- **Proportion** is automatically selected beside **Quantity**
- Beside Level of interest, **select Yes** because we are interested in the proportion of people who said "Yes" to the smoking 100 cigarettes question
- Type in your sample size, use **80** for this example
- Click **Record my choices**.

Next, make your selections in the middle part of the window under **Sampling**.

- **Select 1** repetition
- Click **Go**.



A sample of 80 from the NHANES-1000 population will be selected and drop down onto the middle graph labelled **Sample**.



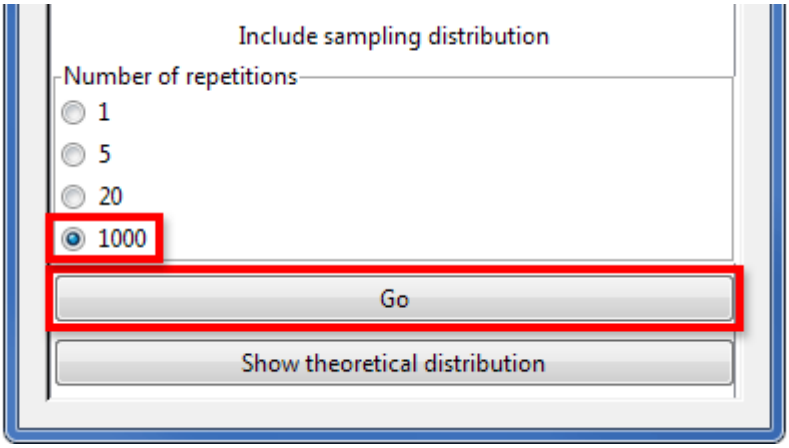
What is happening in the graphics window?

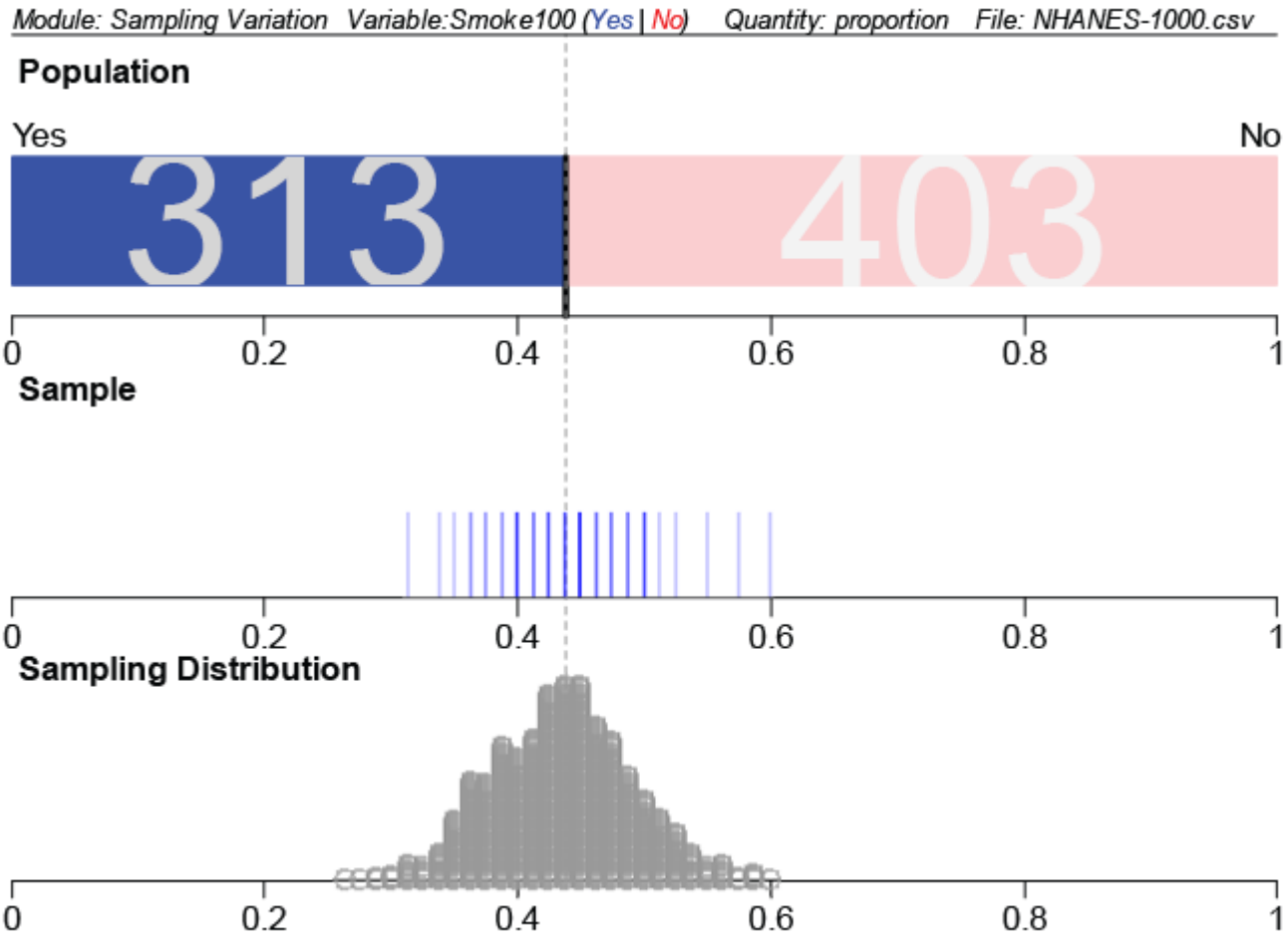
iNZight VIT has taken a sample of size 80 from the population of 1000 and calculated the proportion of people in the sample who have smoked 100 cigarettes or more. In the sample above, the proportion of smokers is about 0.44 or 44% but yours will probably be different. The blue segment is the proportion of smokers in the sample. Remember that our proportion of 'smokers' in the population is in the top plot.

- **Select 5** repetitions and observe 5 more samples of size 80.

Now we'll use the bottom part of the **Analyse** window that says **Include sampling distribution**.

- Under Number of repetitions **Select 1000** and click **Go**.





What is happening in the graphics window?

You should see 1000 different samples of size 80 taken from the NHANES-1000 population. 1000 different sample proportions of 'smokers' are calculated from

the 1000 different samples. Each sample proportion is shown as a blue line in the middle plot and circles in the **Sampling Distribution** plot.

The dotted line is the population proportion.

You will see a sampling distribution of the proportions for 1000 different samples taken from the population build up in the bottom plot.

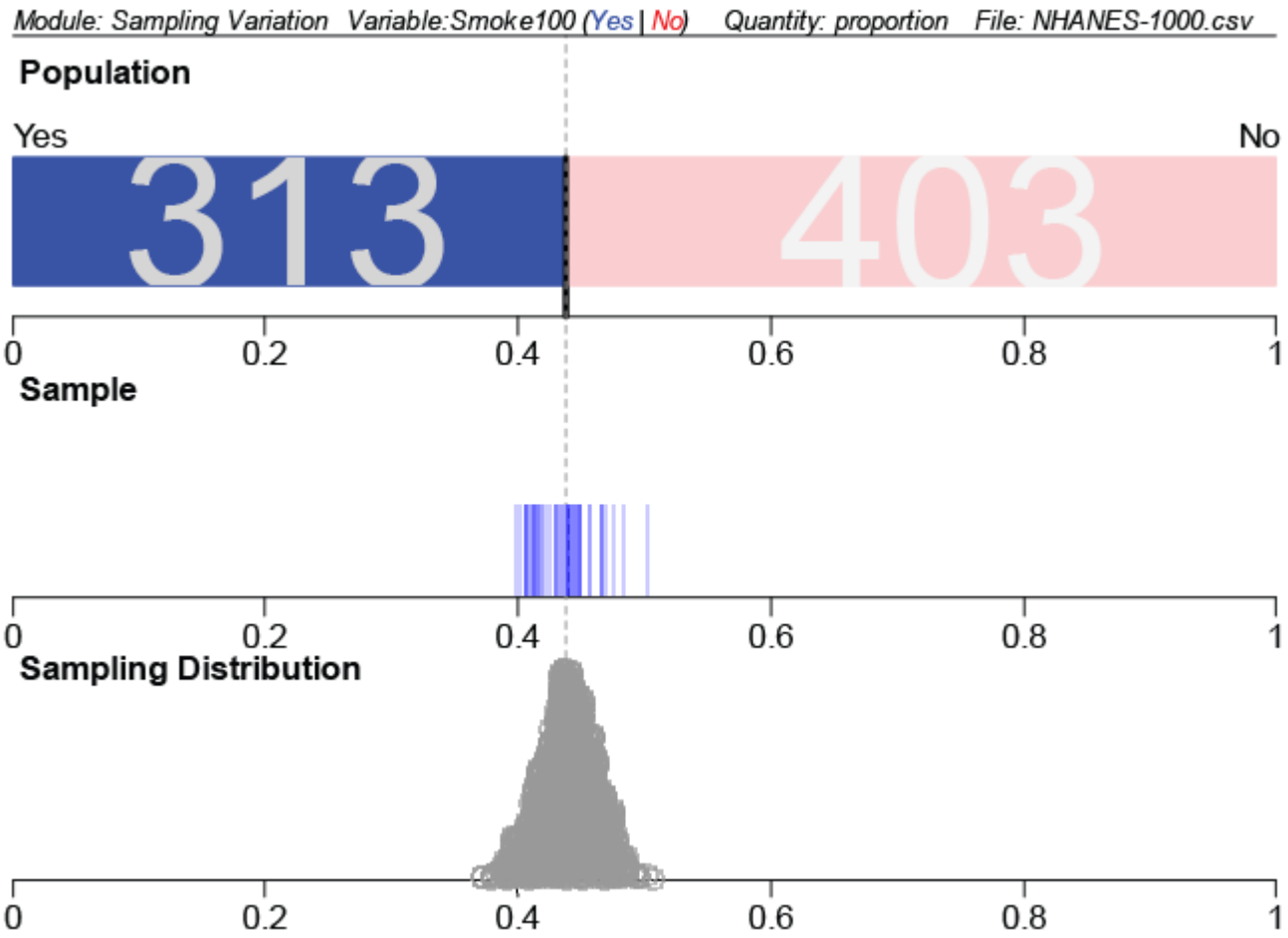
This exercise demonstrates that if we take a sample of 80 from a population of size 1000, our estimate of the proportion of people who have smoked more than 100 cigarettes is likely to be somewhere between about 0.25 and 0.6.

Your sampling distribution will not look exactly like this one but it should be similar.

Exploring the effect of increasing the sample size

With iNZight VIT we can easily explore the effects of changing the size of our samples.

- Change your **Sample size** to **320** (4 times the previous amount)
- Click **Record my choices**
- Under the heading **Include sampling distribution**, click **Go**.



What is happening in the graphics window?

1000 different samples of size 320 are being taken from the NHANES-1000 population. The sample proportion of each of those samples is being recorded as a blue line on the **Sample plot** and a small circle in the **Sampling Distribution**.

PRACTICE (~5 min)

Explore the effects of using different sample sizes and try out some other categorical variables of interest. If you notice or discover anything interesting, post a comment.

Common questions

When I change the sample size and click Go, the sample size changes back to the old value.

You need to click **Record my choices** to make VIT save your new sample size.

Does it matter which Go button I press?

Yes it does. The first **Go** button sets off animations that only happen in the middle graphics panel. The second **Go** button causes the sampling distribution to be formed in the bottom graphics panel. It will use the number of repetitions selected above the button that you press.

What happens if I keep pressing the Go button?

VIT will sample again and add the new samples to the current plot. The exception is when 1000 repetitions are selected. In that case the old display is cleared before starting.

How do I clear my plot and start again?

Click **Record my choices** to clear the plot window and save the current sample size.